Influence of Perspective on Dynamic Tasks in Virtual Reality

Naval Bhandari* Eamonn O'Neill†

Department of Computer Science, University of Bath, UK

ABSTRACT

Users are increasingly able to move around and perform tasks in virtual environments (VEs). Such movements and tasks are typically represented in a VE using either a first-person perspective (1PP) or a third-person perspective (3PP). In Virtual Reality (VR), 1PP is almost universally used. 3PP can be represented as either egocentric or allocentric. However, there is little empirical evidence about which view may be better suited to dynamic tasks in particular. This paper compares the use of 1PP, egocentric 3PP and allocentric 3PP for dynamic tasks in VR. Our results indicate that 1PP provides the best spatial perception and performance across several dynamic tasks. This advantage is less pronounced as the task becomes more dynamic.

Keywords: Dynamic task, virtual reality, first-person perspective, third-person perspective, egocentric, allocentric, virtual environment, dynamic view, immersion, presence, spatial perception, task performance.

1 INTRODUCTION

A dynamic task is one in which the actor's performance of the task is influenced directly by the changing environment in which the task is performed [7, 23, 24, 32]. Edwards [18] states that a task is dynamic when "the environment in which the decision set [to perform a task in a particular way] may be changing, either as a function of the sequence of decisions, or independently of them, or both". In some cases, the factors influencing task performance may be external to the actor, e.g. other cars on the road influencing a driver's performance of a manoeuvre. In other cases, the influences may be intrinsic to the person performing the task, e.g. how tired the driver is.

In contrast, a static task is one where there are no such factors influencing the performance of the task. Such tasks are often deterministic and will arrive at the same result if given the same input [22, 36, 54]. Trivially, using a light switch can be an example of such a task. It can be turned on or off. The same amount of movement of the switch is used each time to turn it off, and vice versa. In normal use there are no factors external to the light switching task which affect this. Combining a static task with an external factor can turn it into a dynamic task. If two valves control the flow of hot and cold water into a container and there is a temperature requirement for the full body of water, the valve turning becomes a dynamic task since the flows of both the hot and cold water need to be changed over time, influenced by the overall water temperature and the flow through the other valve.

Tasks which require substantial human physical movement are typically dynamic. Thus, many elements of sports involve dynamic tasks. Depending on the sport, these elements may include features of the player, of opponents, of equipment, of the performance space etc.

*navs92@hotmail.com
†e.oneill@bath.ac.uk

Previous studies have looked at sports as a form of gaming within VR, but not necessarily the performance or spatial perception of the player during the activity [2, 4, 17, 21, 29, 38, 55]. Other studies have investigated task performance within VR, and have compared it across different user perspectives [1, 20, 46, 47, 53]. However, most studies focus on a single task or do not vary how dynamic the tasks are. Therefore, it is difficult to draw general conclusions on how performance may be affected by how dynamic the task is [22].

Several studies have shown that third-person perspective (3PP) is preferred over first-person perspective (1PP) in games played on a 2D screen [3, 11, 45, 47], although some studies contradict this [12, 15, 16]. In headset-based virtual reality (VR), as the camera is controlled by the player via their head movements, some 3PP paradigms in traditional games are not applicable. However, the relative merits of 1PP and 3PP in VR remain an active research question. Previous research has focused on two representations of 3PP: egocentric and allocentric [44, 49, 52]. In this paper we compare the use of 1PP, egocentric 3PP and allocentric 3PP for increasingly dynamic tasks in headset-based VR, hereafter simply VR for brevity.

We take an empirical approach in this study which allows us to generalise our results concerning performance and spatial perception metrics [35, 41, 43] for dynamic tasks in VR. In Section 2, we identify that a limitation of previous work in this field is the lack of statistical power to substantiate the claims made by some studies. We also note that a large amount of work solely focuses on qualitative measurements, which by themselves cannot confirm their hypotheses. The study in this paper builds on previous studies, combining qualitative and quantitative assessments, and has sufficient statistical power to confirm or reject the hypotheses made.

In this paper, we take three sports as scenarios in which to investigate the influence of user perspective on the performance of dynamic tasks in VR. We chose 3 sports - archery, darts, and tennis - which have common elements including the shared goal of hitting a target but vary in how dynamic the task is in meeting that goal. Our study compares the effects of first and third-person perspectives with regard to these dynamic tasks. This is implemented using camera views which follow the user's avatar (egocentric) versus those which are independent of the avatar (allocentric). We found that 1PP is better than 3PP for both performance and spatial perception, and that allocentric 3PP (A3PP) is better for performance and spatial perception than egocentric 3PP (E3PP). Our results also show that as tasks in VR become more dynamic, the differences in performance and spatial perception between perspectives become less pronounced. We find that for moderately dynamic tasks, spatial perception improves when using A3PP compared to 1PP or E3PP.

2 BACKGROUND AND RELATED WORK

Although considerable research has been conducted on perspective and task in VR, there is little consistency across the literature in agreeing key themes and demonstrating replicated findings.

Many of the studies investigating the effects of different perspectives on task performance adopted the paradigm of a real time camera view of the physical world being streamed to the user's eyes, rather than placing the user in an interactive virtual environment (VE). It is therefore unclear the extent to which we can reliably transfer results from these studies to VEs.

For example, Salamin et al. [46] used cameras to implement a 3PP of the physical environment. Users had a camera attached to their back and wore a headset. This allowed the user to view the camera stream to watch themselves performing actions in real time in the third person. Their experiment consisted of several tasks such as navigation, or passing a physical ball between the user and another person. They concluded that most of the participants preferred 3PP, though this is based on questionnaire results from only 8 participants. They found that 3PP was the preferred choice for interacting with moving objects, or for displacement actions, whereas 1PP was preferred for interacting with static objects directly in front of the user.

The camera used by Salamin et al. [46] was not stereo, with a much smaller field of view (FoV) than current head-mounted displays (HMDs), and therefore provided reduced depth cues to the user. As there were no VEs in their experiment, the tasks available to the participants were applicable only to physical scenarios. Although their study provides some intriguing suggestions of the types of tasks which may be suitable for a 3PP VR experience, some questions remain open.

Salamin et al. [47] attempted to improve upon the 3PP approach in [46] by making the user's body translucent in order to avoid occluding their view while performing a task. This was achieved by creating a mask of the user's head, and using an additional camera to fill in the occluded area. This gave users awareness of their own body while allowing them to see through and beyond it. Salamin et al. found that this approach offered improved performance over traditional 3PP, and was better for dynamic tasks than 1PP. Though they followed a counterbalanced experimental protocol with 24 participants, they reported no analytical statistics on their data, comparing only averages. Participants, on average, preferred the improved version of 3PP for every task.

Salamin et al. [47] investigated several tasks: navigating a room, opening a door, dropping a ball in a cup, and kicking and throwing a ball between users. Most of their results were not presented for analysis but they suggested that 3PP, and their improved version of 3PP, were better than 1PP for dynamic tasks. Though their improved version of 3PP was preferred by users over 1PP and occluded 3PP, both conventional 3PP and the improved version had worse performance than 1PP. This was based on the amount of balls caught/touched by the participant, and the number of "well sent" balls to the experimenter in both a football and a basketball scenario. This result was in contrast to the preferred perspective of the participants.

As with Salamin et al. [47], Englund [20] used a physical camera. Unlike Salamin et al. [47], Englund used a camera controlled by an additional person. As the camera was controlled by a separate person, the camera view was subject to stochastic error due to how they moved the camera to follow the user. Englund [20] compared the navigational performance of users with three different types of view: birds-eye, side-on, and blinded. The birds-eye view had the camera above the participant, pointing directly down towards them. The side-on view had the camera beside the participant, pointing at them. When participants were blinded, they had no visual representation of the scene and had to rely on their other senses. Englund [20] found that participants navigated through an obstacle course quicker using the birds-eye view than the side-on view. Participants finished marginally faster in birds-eye than in side-on view. No comparisons were made between the blinded view and the 3PP conditions.

Kosch et al. [33] created a system which automatically determined the optimal position of the camera above a user to generate a 3PP view. This was similar to the set up in Salamin et al. [46], in which the camera was attached behind the user. The system was tested with a variety of horizontal and vertical camera positions, all pointing towards the participant's head. They concluded that the top-right side of the user behind their head was the preferred position for the camera, based on a qualitative evaluation. Their system was not truly automatic but was manipulated externally by the experimenter. They compared the participant manually adjusting the camera against the experimenter adjusting it. The task was to navigate through a maze. They found that participants performed better with the experimenter adjusting the position of the camera.

Their results showed no significant difference in speed of completion between the conditions. They did find a significant difference in the number of errors made, with fewer made when the experimenter adjusted the camera. However, as the experimenter was manipulating the camera, unexpected movements by the user were not accounted for, requiring them to pause the experiment and readjust the viewing angles. They did not discuss the approach which would be taken to determine the best position and viewing angle of the camera if it were to be a truly automatic system. They suggested that using an Unmanned Aerial Vehicle (UAV) would be a marked improvement, as it would eliminate the need to attach the 3PP camera to the user. Developing a UAV with the capability to determine the best viewing position while adjusting for even tiny head movements would, however, be challenging.

Studies have also investigated the role of user perspective in tasks within fully virtual environments rather than streaming video of the physical environment to the user's display. Although in this case the camera is virtual rather than physical, there remain inconsistencies across these studies, e.g. in their use of traditional displays or HMDs, their experimental conditions, and their findings.

Alonso et al. [1] compared catching a virtual ball in VR in 1PP with different forms of 3PP. They found that the number of balls caught decreased significantly as the camera moved farther away from an egocentric point of view. They moved the camera back in intervals of 0.75m, with the height constant at 2m. As the camera was moving backwards in one axis instead of diagonally, this could have had an adverse effect on performance. They concluded that 3PP should not be used for VR tasks unless aligned as closely as possible to 1PP. No avatar was rendered during Alonso et al.'s [1] experiment. Rather, just a single controller was rendered. The lack of an avatar could have contributed to the performance degradation as the camera was moved farther away from an egocentric 1PP. Most video games which involve the personification of a character in 3PP use a representation of that character.

Brunye et al. [9] compared two egocentric perspectives for navigation tasks in a VE. Participants watched a simulated video feed from either the front of a vehicle (route), on top of one (survey), a combination of both, or neither. Brunye et al. tested how efficiently participants would navigate through the VE once exposed to these feeds. They also tested the diminishing effect that each perspective had as participants increased their exposure to the feeds after repeated trials. The route perspective is similar to 1PP, and survey to 3PP. They found that the route perspective was optimal for navigating short distances, particularly in straight lines, whereas the survey perspective was better for long distances.

Debarba et al. [13, 14] presented two studies which focused on how embodied users felt in VR under 1PP and 3PP. The first task presented to participants was to reach out to an area. The second task had participants walk onto a ramp surrounded by wooden flooring. The wooden flooring then dropped (in the VE), revealing a pit below the participant's virtual avatar. They used full-body tracking to allow for accurate control of the avatar. The first task had three conditions: 1PP, 3PP, and allowing participants to alternate between 1PP and 3PP. While participants felt safer in 3PP for the second task, they preferred 1PP for performing the task. Interestingly, they found that allowing participants to alternate between perspectives did not break the feeling of body ownership and embodiment.

Medeiros et al. [37] compared 1PP and 3PP across different user avatar representations for navigational tasks. Their design implementation of 3PP was based on the study by Kosch et al. [33], which is an egocentric spatial representation. They designed 3 representations of the user: a block-like avatar, a human-like virtual avatar, and a point-cloud mesh representation of themselves. For the last representation, they captured a point-cloud mesh of the user's physical body, and inserted this representation into the VE. The tasks presented to participants were based around how users interact with obstacles within a small space instead of navigating through a larger area. They found that 1PP was the most efficient and effective perspective for navigation tasks. They stated that users felt a higher level of embodiment in 1PP than in 3PP for all representations except the point-cloud mesh, where there were no differences. These results suggest that a realistic representation of the user is important for eliciting full-body ownership of the avatar when in 3PP.

Gorisse et al. [25] explored different perspectives in immersive VEs by testing user performance on object perception and navigation. They compared 1PP against 3PP, measuring the time taken to navigate an area, and how accurately participants interacted with items in a scene, with the user tracked wearing a motion capture suit. They found that while the user's spatial perception improved in 3PP, they performed equally well on the task of deflecting projectiles. They concluded that 1PP was better for embodying a virtual body, whereas 3PP was better for spatial awareness and perception. This was supported by a statistically significant difference in perception delay between 1PP and 3PP. 1PP was also found to have offered a significant advantage during navigation tasks without obstacles and with obstacles. This contrasted with the findings of Salamin et al. [47, 48] who suggested that 3PP was advantageous when performing dynamic tasks, such as navigation.

Török et al. [49] compared 1PP, 3PP, and an aerial view. While 1PP could be used only in an egocentric view, the other two conditions were tested in both an egocentric and allocentric version. The egocentric version of 3PP oriented the camera based on which way the avatar was pointing in the VE. The allocentric version faced true north, regardless of where the avatar was pointing. The task was to search and retrieve objects. While they found no significant performance difference between 1PP and 3PP, participants were significantly faster at performing the task in E3PP than in A3PP. The experiment was conducted using a tablet display so the results are not directly applicable to headset-based VR applications, but it does again provide intriguing suggestions of the different effects of egocentric and allocentric reference frames.

The lack of consistency across previous studies in many aspects, from the control of physical and virtual cameras for generating 3PP to the tasks performed, makes it difficult to compare findings. There are also variations even in the definition of dynamic tasks across previous studies. For example, Salamin et al. [47] stated that non-static external objects introduced into the task make it dynamic. Examples given by Salamin et al. were passing a ball between the participant's and experimenter's feet, or catching a ball. This implied that a response element is required for a dynamic task. Whereas Johnson et al. [30] claimed that movement from the participant alone makes a task dynamic. In this case, the movement making it dynamic was physically moving away from a work area to a separate one, or retrieving static objects to a work station. Gorisse et al. [25] used two tasks which fit with these two definitions respectively. The first task involved dodging spheres, and the second involved navigating through an environment and interacting with a static object.

While there have been studies comparing first and third person views, none has compared them across tasks with varying levels of dynamic elements. To the best of our knowledge, there are also no studies which have compared an egocentric 3PP against an allocentric 3PP within immersive headset-based VR. Much of the previous work comparing 1PP and 3PP has relied on small scale qualitative data for its findings, with an unknown relationship between a user's performance and their stated preferences.

3 Experimental Design

In our study, we compared the use of 1PP, egocentric 3PP and allocentric 3PP for varying dynamic tasks in VR. This study combined experimental trials using different perspectives and dynamic tasks with demographic and user experience questionnaires. The resulting collection of quantitative and qualitative data allowed us to analyse how different perspectives and dynamic tasks affected the performance, spatial perception and preferences of users. While Johnson et al. [30] claim that movement of the user makes the task dynamic, Salamin et al. [47] state that non-static external objects make the task dynamic. We incorporated both of these definitions in all of our tasks, increasing the dynamic elements across our conditions.

We presented user perspective in three distinct views:

- First-person perspective - standard view in a VR HMD;
- Egocentric third-person perspective - viewing an avatar (representing the user and mapped in real time to the user's physical movements) from behind through an HMD, where the camera view is kept constant relative to the avatar's head;
- Allocentric third-person perspective - viewing an avatar (representing the user and mapped in real time to the user's physical movements) from behind through an HMD, where the camera is independent of the avatar's head.

To investigate the influence of these perspectives on the performance of dynamic tasks in VR, we chose 3 sports which, while they have common elements, have different mechanics and varying dynamics: archery, darts, and tennis. Archery is the least dynamic of the three, involving a steady pose with built up potential energy. Once the participant has drawn back the arrow and lined up the shot, all they need to do is release it by softening their grip on the VR controller. Although there are very small movements of their hands and head which may impact their shot, it is a largely static task. Darts is more dynamic than archery, with more movement by the participant, in particular wrist, finger and arm movement. The tennis task is the most dynamic of the three, involving diverse factors including where an opponent hits the ball, the position and other characteristics of the player, and the position and movement of the racket as it hits the ball.

3.1 Hypotheses

Despite the challenges in comparing the relevant literature, some key findings do seem, at least tentatively, to emerge from previous work, and we aimed to build on them in this study. Several previous studies have concluded that 1PP is the preferred choice of users in most scenarios [1, 25, 33]. However, these studies are largely reliant on qualitative user feedback rather than quantitative performance metrics, and user preference does not always align with user performance measures. Hence we set out to test this hypothesis:

H1: Giving a user a first-person perspective (IV) will result in better performance (DV) compared to any third-person perspective.

Spatial perception is a measurement of a user's ability to achieve an accurate understanding of spatial relationships and characteristics within the VE [6, 26, 28]. This includes a user's ability to judge distances and take account of spatial transformations. Gorisse et al. [25] found that spatial perception was significantly better in 3PP compared to 1PP, however, the metrics used to determine this may be subject to some uncertainty. The metric for detecting perception of an object was determined by when the object entered the HMD's viewing frustum. This is a large area, and will mathematically always contain 1PP's viewing frustum inside the 3PP alternative. If all other conditions remain the same, the viewing frustum of 3PP will always contain an additional zero or more objects compared to the viewing frustum of 1PP. Hence, we tested a second hypothesis:

H2: Giving a user any third-person perspective (IV) will result in better spatial perception (DV) compared to a first-person perspective.

The aim of all the tasks in our experiment is to hit a target. Hence, we used metrics for spatial perception related closely to that aim. These included, for example, the accuracy of hitting the target, and the 3D transformation deviation per attempt at hitting the target. See below for more details of the metrics used.

Török et al. [49] concluded that, although there were no significant differences between 1PP and 3PP, egocentric 3PP was significantly better than allocentric 3PP for both spatial and temporal performance on a navigation task. Navigation performance has also been shown as a measure for spatial perception [10]. Hence, our third hypothesis is:

H3: Giving a user an egocentric third-person perspective (IV) will result in better task performance (DV) and spatial perception (DV) compared to an allocentric third-person perspective.

3.2 Participants

Data were collected from 30 participants (17 males, 13 females; age 18-56, mean 26). They were a mixture of university students and staff. 14 had normal vision, and 16 had corrected to normal vision. 7 had no VR experience, 14 had a little, 5 had a moderate amount, and 4 had a lot of experience in VR. There were 27 trials per participant, per sport. All data were processed using Statistical Package for the Social Sciences (SPSS) and R [42].

3.3 Procedure

All participants took part in this study on the basis of written and informed consent. They were free to opt out of the study at any time and without delay. The participants were not reimbursed with monetary payment for their time, nor did they receive any other form of reward for their participation.

Our data collection took three forms: we collected demographic data on the participants via a pre-experiment questionnaire; we ran a series of trials with each participant to collect performance and spatial perception data; and we used a questionnaire after each condition to collect data on the user's experience.

The experimental VE was developed in Unity 3D[1]. All physical parameters were set to the default using the Unity physics engine. The experiment ran on an Alienware R15 laptop, with an i7-6820HK CPU, and a Nvidia GeForce GTX 980M GPU, with 16GB DDR3 RAM. The HMD used was the *HTC Vive*[2]. The HMD allows a room-scale environment, where the user's head and controllers are fully tracked with 6 degrees of freedom (DOF) to an accuracy of 0.3mm in all directions [5, 34, 39]. The tracking environment was a 3m x 3m space.

We used an avatar to depict the participant in the rendered VE. The avatar moved based on inverse kinematics (IK). We tracked the user's head and two controllers, one in each hand. These were mapped to the head and respective hands of the avatar. The rest of the avatar body moved based on these three tracked objects. As only small movements were made by participants, we had no need to track their feet. We initially planted the feet of the avatar on the floor, and if the participant moved too far in one direction, the avatar's body would adjust to the tracked hands and head.

For E3PP, a virtual camera was placed behind the avatar's body, focused on their head position. This let participants see a one-to-one mapping between their movements and the avatar's movements in real time. We offset the camera to 50cm behind and 25cm above the avatar's head. This gave us a linear distance of 56cm from the head position to behind the avatar's head. This is in line with Alonso et al. [1], who found that minimising the distance between the user and the camera maximised the user's performance. We calculated the orientation of the camera as:

[1] https://unity.com/

[2] https://www.vive.com/uk/

$$\mathbf{f} = (\widehat{\mathbf{s} - \mathbf{t}}), \tag{1}$$

$$\mathbf{r} = (0, 1, 0) \times f, \tag{2}$$

$$\mathbf{u} = f \times r, \tag{3}$$

$$R = \begin{bmatrix} \mathbf{r} & 0 \\ \mathbf{u} & 0 \\ \mathbf{f} & 0 \\ \mathbf{s} & 1 \end{bmatrix}, \tag{4}$$

where **s** and **t** are the camera and head position respectively. $(\widehat{\mathbf{s} - \mathbf{t}})$ is the normalisation of the vector. R is the resulting transformation matrix applied to **s** to rotate towards the head position. This ensures that no matter how the participant moves, the camera will always be pointing centrally towards the head of the participant.

For A3PP, the placement of the camera at the beginning of each trial was behind and above the avatar in exactly the same position as for E3PP. As the user moved, the camera did not move relative to the avatar but independently of it. As the participant moved their head and hands, the camera and avatar would make the same movements. This meant, for example, that if the participant turned through 180°, both the camera and avatar would also turn through 180°, so that the camera is now pointing away from the avatar, and the user therefore no longer has the avatar in view (illustrated in Figure 1).

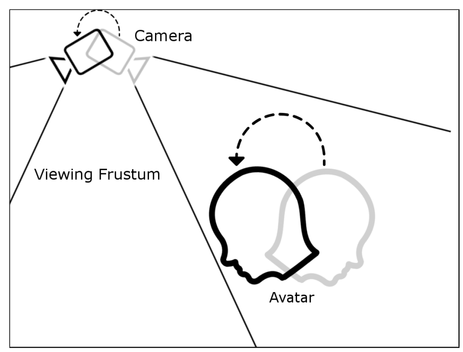


Figure 1: Illustration of head movement in A3PP

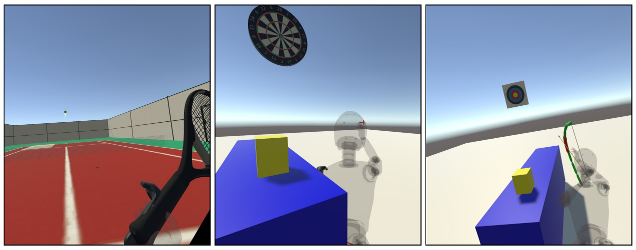
The avatar was translucent to provide a balance between allowing the participant to see their own movements and not occluding what is in front of the avatar. We chose this approach following Salamin et al. [47] who found that both performance and user preference were worse for a fully opaque avatar as opposed to a translucent one.

Our experiment consisted of three sports related dynamic tasks: archery, darts, and tennis. Each task followed a similar protocol. The participants were given initial practice prior to starting the task for each condition.

Participants were given a virtual bow and unlimited arrows for the archery task. The participant had to grab an arrow from a box beside them, before nocking it on the bowstring. For both archery and darts, we created a 3x3 grid along the x-z plane. As practice, participants had to hit 5 targets, prior to the task beginning and data being collected for each condition. Each time they hit a target, it spawned in a new location on the grid. After the initial practice, participants had to hit the target in all 9 different locations on the grid.

In the archery task, participants had to shoot a standard circular archery target, shown in Figure 2. The target spawned randomly on the grid, and each time the participant hit it, it spawned in a new location. The order of targets was random for each trial. This was repeated for each of the 3 perspectives. The bow could be switched between controllers to account for the participant's dominant hand. Participants had to attach the arrow to the bow, and pull the arrow to shoot it. They aimed the bow using one controller, and adjusted the bow tension with the other.

The darts task was similar to the archery task. Participants had to pick up a dart with their dominant hand, and use the movement of their hand to throw the dart to the target. Participants had the same initial 5 practice trials, and subsequently 9 targets to hit after the practice. The grid for the darts task was placed at half the distance of the archery task.



(a) 1PP tennis (b) E3PP darts (c) A3PP archery

Figure 2: Different perspectives in each sport

In the tennis task, participants had to hit a ball, which came at them from the other end of the court, to a glowing square target which covered a 1x3 grid on the court. The targets for the participant to aim for were the left, right, and centre areas at the opposite end of the court. We randomised the starting position of the ball across the same 3 areas. This gave us 9 combinations of where the ball came from and where the target was. The participants had a virtual tennis racket attached to their hand, which could be switched between controllers to allow for their dominant hand. Participants had to hit the ball coming towards them to the target with the racket.

Participants were given 15 practice trials for the tennis task per condition before data were collected. A trial was an individual ball shot. After the 15 practice trials were completed, participants had to hit the ball to each of the 9 combinations of ball source position and target position. Each ball appeared 3 seconds after the previous one. To avoid interference with other balls, we destroyed each ball from the scene after 6 seconds.

From pilot studies, we determined that the tennis task was difficult; when hit, the ball seldom went towards the target. In order to avoid a floor effect due to this observed difficulty in hitting the virtual tennis ball to the target, we introduced a bias towards the target. This bias consisted of an 86% linear interpolation of the desired vector to hit the target and the actual vector produced from the participant hitting the ball. Further testing gave an average number of targets hit of 5.37 from a total of 9 per condition and we therefore concluded that a floor effect would not occur in our experimental trials. This was borne out by the results. In addition, rather than require the participant to continue hitting balls until they successfully hit the target, which might have been a very long time for some participants, we limited the attempts to 15 balls per target. This meant that a trial could finish without the participant hitting the target.

Several VR tennis games are available[3] [4]. They currently have three types of mechanics for moving around the court. Players can either move according to their physical movements, automatically teleport to the correct location, or accelerate automatically towards the ball. There are pros and cons with each of these, with reviews of the games suggesting that most players opt for automatic teleportation. Manual movement gives little scope to move around, as most VR users are playing in spaces much smaller than a tennis court, while automatic acceleration tends to give users motion sickness.

Because of the potential issues with player movement, we constrained the position of the participant to an area in the centre of the court. The ball was shot towards the racket from the point of spawning so that participants never had to move far from their current position.

There were 9 targets per condition, per sport. This gave us a total of 27 trials per sport, and 81 per participant. After each condition, the participant was asked a series of questions about that condition.

The dependent variables (DVs) measured were:

- The performance of the participant. The performance metrics for archery and darts were:
  - Time to hit target
  - Number of attempts to hit the target

  The performance metrics for tennis were:
  - Time to hit target
  - Number of shots
  - Number of successfully hit shots
  - Number of bounces
  - Number of successfully hit targets
- The spatial perception of the participant. As noted above, spatial perception includes a user's ability to judge distances and take account of spatial transformations. In the context of our experimental conditions, that entails effectively taking account of the distances and transformations required by the 3 dynamic tasks. The spatial perception metrics for archery and darts were:
  - Distance to centre of target when hit
  - Velocity of arrow/dart when shot/thrown
  - Rotational consistency of arrow/dart when shot/thrown

  The spatial perception metrics for tennis were:
  - Velocity of ball
  - Rotational consistency of ball
  - Distance to centre of the racket
  - Distance to centre of target when hit
  - Distance to centre of target when missed

Rotational consistency was calculated as the standard deviation of the angular magnitude between the shots taken by the participant.

We employed a related samples design with randomisation to avoid bias for any particular IV and order effect. Each participant undertook every condition in a random order. We calculated the minimum sample size as 20 using a power test with a minimum power of 0.8 [51], given an effect size of 0.4 (Cohen's D).

4 Results

4.1 Performance and spatial perception

One-way repeated measures ANOVA were conducted to analyse the effects of perspective on the performance and spatial perception data captured during the experiment. Post-hoc Tukey HSD (honestly significant difference) tests were conducted for pairwise comparisons between IVs.

For the archery task, the main effects that we observed were statistically significant differences in task performance based on the number of shots (**AS**) ($F(1,29) = 10.137$, $p = .004$), and the time (**AT**) taken to hit the target ($F(1,29) = 6.834$, $p = .014$). The mean number of shots taken in 1PP was 45% less than in E3PP, and 35%

[3] Dream Match Tennis VR PS4
[4] First Person Tennis - The Real Tennis Simulator

less than A3PP. The average time taken to hit the target was 9.1 seconds in 1PP, 14.3 seconds in E3PP, and 13 seconds in A3PP.

All the performance metrics support **H1**, while we have mixed results for **H3**. Participants in E3PP were shown to be significantly worse in both the number of shots taken ($p = .051$), and the time taken to hit the target ($p = .05$) than in A3PP.

We found a significant effect on the accuracy (**AA**) of the shot when hitting the target ($F(1,29) = 7.455$, $p = .011$), and on the rotational consistency of the arrow (**AR**) as it was shot ($F(1,29) = 10.137$, $p = .004$). The average accuracy of a shot in 1PP was 27.4cm away from the centre, compared to 31.1cm in E3PP, and 33.3cm in A3PP. There was a large increase in rotational inconsistency from 1PP to both E3PP (45%) and A3PP (34%). These results lead us to reject **H2** and provide mixed results for **H3**. While the rotational inconsistency was worse for E3PP compared to A3PP, the accuracy was higher.

For darts, we observed a significant difference between perspectives for the number of shots taken (**DS**) ($F(1,29) = 6.063$, $p = .021$), the time taken (**DT**) to hit the target ($F(1,29) = 6.246$, $p = .019$), the average accuracy (**DA**) of the shot when hit ($F(1,29) = 5.147$, $p = .032$), the rotational consistency of the dart (**DR**) as it was released ($F(1,29) = 6.063$, $p = .021$), and the velocity of the dart (**DVe**) as it was released ($F(1,29) = 4.234$, $p = .05$).

The mean number of shots taken was 18.9 in 1PP, 31.2 in E3PP, and 25.6 in A3PP. The average time to hit the target was 7.86 seconds in 1PP, 14.93 seconds in E3PP, and 11.8 seconds in A3PP. The average accuracy of hitting a target was 35.9cm in 1PP, 38cm in E3PP, and 32.8cm in A3PP. The rotational consistency in A3PP was 6.1% higher than in 1PP, and 46% higher than in E3PP. The average release velocity was 8.15 m/s in 1PP, 11.5 m/s in E3PP, and 9.75 m/s in A3PP.

These results support **H1**, and partially support **H2**. While participants were faster and took fewer shots in 1PP, they were more accurate in A3PP ($p = .05$). A3PP also outperformed E3PP with regard to shots taken ($p = .048$), time taken ($p = .039$), and accuracy ($p = .019$), thus leading us to reject **H3**.

The only significant effect found in the tennis task was on the distance from the centre of the racket (**TRA**) ($F(1,29) = 4.441$, $p = .045$). The average distance was 10.4cm in 1PP, 9.5cm in E3PP, and 9cm in A3PP. The distance was shorter for A3PP than for E3PP ($p = .059$), which in turn was shorter than for 1PP ($p = .047$). This result supports **H2**, and leads us to reject **H3**.

DV	**1PP**		**E3PP**		**A3PP**		
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	p
AS	12.4	6.94	18.0	14.4	16.7	7.58	.004
AT	9.10	9.77	14.3	13.2	13.0	8.78	.014
AA	.274	.099	.312	.098	.333	.083	.011
AR	1.38	.771	2.00	1.60	1.86	.842	.004
DS	18.9	9.72	31.2	25.9	25.6	17.8	.021
DT	7.86	4.98	14.9	14.2	11.8	10.3	.019
DA	.359	.101	.381	.072	.328	.088	.032
DR	2.10	1.08	3.47	2.88	2.85	1.98	.021
DVe	8.15	3.64	11.5	7.50	9.75	5.32	.050
TRA	.208	.040	.191	.043	.184	.044	.045

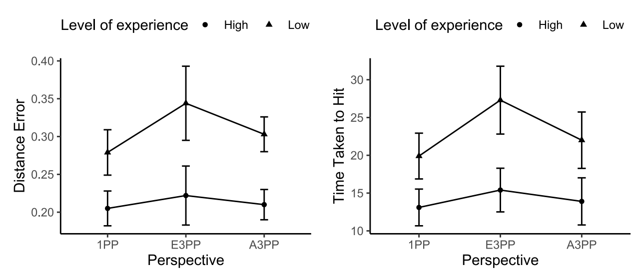
Table 1: Means, standard deviations and ANOVA results for performance and spatial perception data

We observed several interaction effects between perspective and sport, i.e. between perspective and how dynamic the task is. As the sports had similar metrics but each metric was not directly comparable, we compared them using a MANOVA. We found that the number of shots taken increased from archery to darts to tennis ($F(1,89) = 27.62$, $p =< .001$). However, we found that the difference in the number of shots between perspectives became smaller as the tasks became increasingly dynamic. This decrease in the differences between perspectives was also found for the time taken to hit the target ($F(1,89) = 12.36$, $p < .001$), and the accuracy of the shot ($F(1,89) = 23.69$, $p < .001$).

4.2 Demographic Data

We conducted ANCOVA and MANCOVA tests on our demographic data, and saw a mixture of results when correlating with the performance metrics. Some studies have shown differences between male and female spatial perception [40] and performance [27]. We found that male participants hit targets significantly more quickly in darts ($F(1,29) = 4.32$, $p = .047$) and tennis ($F(1,29) = 8.71$, $p = .006$), but were less accurate ($F(1,29) = 7.49$, $p = .011$) than females. They also required more shots in all the sports ($F(1,89) = 11.23$, $p < .001$). Women were more accurate ($F(1,29) = 11.3$, $p < .001$) and faster ($F(1,29) = 8.11$, $p = .008$) at hitting targets in archery than men. Women were consistently more accurate in each sport. While women were faster than men at hitting targets in archery, they were slightly slower than men in darts, and much slower in tennis. As the tasks became more dynamic, both men and women took longer to hit the targets but women's speed declined more.

We also saw trends with respect to vision and occupation. Participants who had normal vision were consistently more accurate ($F(1,29) = 3.98$, $p = .048$) and slower ($F(1,29) = 9.49$, $p = .005$) than participants with corrected vision in both archery and darts. Students were faster ($F(1,89) = 4.97$, $p = .002$) and more accurate ($F(1,89) = 3.93$, $p = .008$) than university staff across all tasks, although this may well be confounded with factors such as age and experience in VR.



(a) Average distance error

(b) Time to hit target

Figure 3: Performance across tasks for different levels of experience in third-person games

We asked participants how much experience they had with third-person games, from no experience at all to a great deal of experience. 14 participants had moderate or greater experience, and 16 had little to no experience. Participants who had moderate or greater experience with third-person games outperformed those with little to no experience across all tasks ($F(1,179) = 61.2$, $p < .001$). In contrast, as Figure 3 illustrates, participants who had moderate or greater experience with third-person games were not significantly more accurate ($F(1,89) = 2.43$, $p = .130$) or quicker ($F(1,89) = 2.23$, $p = .145$) in any perspective. We also found that participants who had high experience in first-person games were significantly better than those who had low experience ($F(1,179) = 7.49$, $p = .011$). These results were similar although less pronounced for participants who had experience in first-person games than for those who had experience in third-person games. This suggests that participants who have had exposure to playing third-person games are less affected by the transition from 1PP to 3PP than those have not had such exposure. We also observe that previous experience in video games was significantly more important than previous experience in the equivalent physical sports ($F(1,179) = 8.70$, $p = .006$). We found no differences on any of the metrics between participants with varying levels of experience in real world tennis ($F(1,89) = 1.27$, $p = .269$), darts ($F(1,89) = 1.97$, $p = .172$) or archery ($F(1,89) = .005$, $p = .942$).

4.3 Post-Condition Questionnaire

A questionnaire was presented to each participant after each condition to capture a subjective assessment of each condition. This allowed us to investigate whether the participant's opinions and preferences were congruent with their quantitative performance and spatial perception data. The questionnaire collected data on the participant's sense of presence, comfort, and ease of use [19,31,50]. We categorised the answers to each question into performance or spatial perception, allowing us to compare these results with the quantitative data.

A Likert scale was used for each question, composed of *strongly disagree, disagree, agree,* and *strongly agree*. We assigned discrete integer values from 0 (strongly disagree) to 3 (strongly agree) to these labels for analysis. As our rating scale was not continuous, we conducted multivariate non-parametric related samples Friedman Tests.

We obtained several statistically significant results for the post-condition questionnaire across perspectives. Participants felt that 1PP and A3PP were significantly easier to perform in than E3PP ($\chi^2(2) = 10.11$, $p = .043$). There were no significant differences between 1PP and A3PP. This indicates that users did not find 3PP *per se* to be more difficult but they did find an egocentric version harder to perform in. This pattern was consistent across the level of discomfort participants felt ($\chi^2(2) = 13.98$, $p < .001$), and how easy participants felt it was to hit each target ($\chi^2(2) = 11.01$, $p = .009$).

Participants felt that they needed to learn more things in E3PP than in A3PP ($\chi^2(2) = 11.25$, $p = .004$). This may be attributed to the fact that the camera mapping in E3PP does not directly map to the head movements of participants, and thus requires longer to become accustomed to. Participants felt as if they did better in both 1PP and A3PP, and felt less discomfort than in E3PP. These findings are congruent with our quantitative results which showed that participants performed consistently better in 1PP and A3PP than in E3PP.

Several significant results were observed across the tasks. Participants felt more immersed in archery compared to darts ($\chi^2(2) = 21.46$, $p =< .001$), and more immersed in darts compared to tennis ($\chi^2(2) = 19.2$, $p =< .001$). The same pattern of responses held for participants' desire to stay in the VE ($\chi^2(2) = 10.34$, $p = .006$) and their discomfort levels ($\chi^2(2) = 11.11$, $p = .004$). As the dynamic element of the task increased, participants' preference for it dropped and their discomfort increased. Participants' feelings of how easy the task was decreased from archery to darts to tennis ($\chi^2(2) = 23.69$, $p =< .001$). Participants also found that they needed to learn less to get going with the sport ($\chi^2(2) = 11.08$, $p = .004$) and that the target was easier to hit ($\chi^2(2) = 43.3$, $p =< .001$) in archery than in darts, and in darts than in tennis.

As the questionnaire residual data did not follow a normal distribution, we conducted a non-parametric test for interaction effects between perspective and dynamic task. We used the F1-LD-F2 design [8,9] to conduct a non-parametric version of a two-way repeated measures ANOVA. We performed a post-hoc Tukey HSD test to examine the pairwise factorial contrasts between the IVs.

We observed an interaction effect for how difficult participants found each sport ($F(4) = 2.579$, $p = .038$). Participants found tennis harder than archery when in 1PP or E3PP but not in A3PP. Similar effects were observed for their discomfort levels ($F(4) = 2.558$, $p = .039$). Participants found tennis significantly more discomforting than archery when in E3PP but not in 1PP or A3PP. This suggests that as tasks become more dynamic, the participant's preference for using 1PP or A3PP over E3PP increases. This provides an intriguing contrast with the quantitative performance data which suggest that as tasks become more dynamic, the differences in performance across the different perspectives become smaller.

5 Discussion

We investigated the influence of altering the user's perspective on their performance and spatial perception across varying dynamic tasks in VR, explicitly taking account of the interaction between how dynamic the task is and the perspective of the user. We also investigated the relationship between users' preferred perspective and their performance and spatial perception. We found that as tasks become more dynamic, the effect that user perspective has on performance and spatial perception decreases. 1PP tends to be the optimal choice for static to moderately dynamic tasks in VR, and if 3PP is to be used, an allocentric version is the best choice. Our results also indicate that user preference is congruent with performance.

As noted above, archery is the least dynamic of the three tasks, and is a largely static task. Both the number of shots taken and the time taken were less in 1PP than in either 3PP condition. The accuracy was also better in 1PP. A Pearson product-moment correlation was run to determine the relationship between accuracy and time taken to hit the target. There was a strong, positive correlation between them which was statistically significant ($r = .840$, $n = 30$, $p =< .001$).

We can infer from the results that the consistent viewpoint of 1PP and A3PP facilitated participants hitting the target more quickly and in fewer shots than in E3PP. The rotational and positional movements of the camera following the avatar in E3PP are compounded by the movements made by the participant, resulting in a less consistent viewpoint.

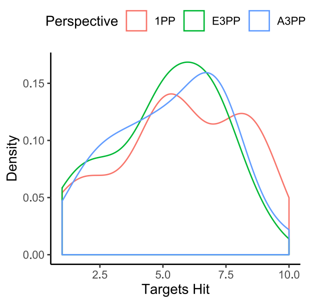
The performance and spatial perception disparities between 1PP and A3PP may be attributable to the different point of view (POV) of the camera. The participant will see the angle of the bow at different POVs, making it harder to align to the target. This meant it took longer to aim the bow. This is also shown when comparing A3PP to E3PP. As the camera takes an egocentric view of the avatar, it is more aligned with the viewpoint in 1PP, particularly with the angle the camera is pointing towards. With A3PP, if participants are looking directly ahead, they align their eyes with where the arrow is pointing.

Similarly to archery, the results for darts favour 1PP in both the number of shots taken and the time required to hit the target. The increase in number of shots in A3PP is 35% higher for both archery and darts compared to 1PP. In E3PP, however, the difference rises to 45% in archery and 65% in darts.

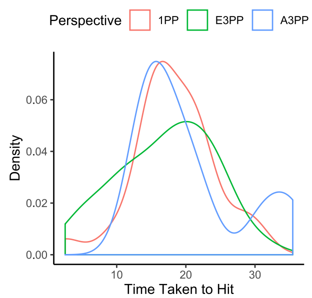
The time taken in A3PP is 50% higher than in 1PP for darts, and 43% higher for archery. The time taken in E3PP is 90% higher than in 1PP for darts, and 57% higher for archery. While darts is more dynamic than the archery task, the differences between 1PP and A3PP remain fairly consistent. Performance and spatial perception in E3PP are shown to degrade with the more dynamic task. As there is more movement by the participant in this task, it is likely that the viewing inconsistency and unnatural movements of E3PP contribute to poorer performance.

For the darts task, accuracy is highest in A3PP. This somewhat contradicts the fact that more shots are required, and it takes longer to hit the target in A3PP than in 1PP. When the target is eventually hit, participants are significantly more accurate in A3PP than in either 1PP or E3PP. As the FoV of the participant is larger, they can see the dart's path more clearly and for longer. 1PP allows for only a narrow view of the dart, as participants can see only the back of the dart as it is shot. Because the camera follows the avatar's head in E3PP, this narrow view of the dart is also present, which may result in a similar effect to 1PP.

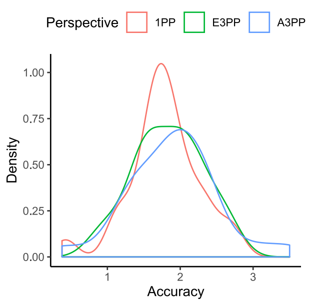
When the participant is tracking the dart, their view of it may perhaps be partially blocked by the avatar even though it is translucent. In A3PP, the dart can be seen from slightly above, and as it moves, the participant does not have to look at the avatar. This may have allowed participants to track where the dart lands with higher precision, allowing for a more accurate shot.



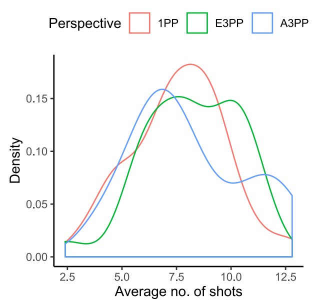
(a) No. of targets hit



(b) Time to hit target



(c) Accuracy



(d) Average no. of shots

Figure 4: Distribution plots for tennis

The tennis task is the most dynamic and most difficult of the three, and many participants missed the ball when attempting to hit it. We observed a significant difference in where the tennis racket struck the ball, measured as the euclidean distance from the centre of the racket. The distance from the centre was the smallest in A3PP, followed by E3PP, then 1PP. As there was a high speed ball coming towards the player, it was difficult to see the ball and respond by hitting it. When in 3PP, the participant could see the ball from farther away. Their racket was also in view, as opposed to 1PP where it sat outside the participant's visual periphery, often to the side of the participant. This allowed the participant to hit the ball more often, and closer to the centre of the racket. Proximity to the centre of the racket is generally regarded in tennis as producing a better shot, and these results suggests that it is easier to hit in 3PP.

Figure 4 illustrates that there were no significant differences between perspectives in performance or spatial perception for tennis, which was the most dynamic task. The distribution plots show that the peaks are in similar positions. We see that variability is higher for E3PP, indicating that participants were more likely to fall on the extreme ends of performance.

A general trend was that as the tasks became more dynamic, they became more difficult, they were enjoyed less, and the perceived discomfort was higher. Participants found it more difficult to hit the target as the task became more dynamic, regardless of perspective.

The quantitative results combined with the results from the post-condition questionnaire suggest that **H3** is rejected, and that an allocentric 3PP is better than an egocentric 3PP in most conditions across the measures of user performance, spatial perception and user preference.

The results are mixed for **H2**, where we reject it for the least dynamic task of archery and accept it for the more dynamic tasks of darts and tennis. This suggests that how dynamic a task is has an influence on how spatial perception is affected by perspective.

The results are also mixed for **H1**, where we reject it for the most dynamic task of tennis and accept it for the less dynamic tasks of archery and darts. This again suggests that how dynamic a task is has an influence on how performance is affected by perspective, but in the opposite direction to H2.

6 Limitations and Future Work

The apparatus used for the sports (i.e. bow, arrow, dart and tennis racket) do not match particularly well the ergonomics of a VR controller, a common usability issue with many VR applications and games. While some matched better than others, this may have had an impact on the participants' performance. This is most notable in the darts task, where the user grips the controller like a tennis racket with the entire hand. They were required to squeeze the controller, and soften their grip to release the dart. This is an unnatural movement quite unlike throwing a physical dart in the real world.

The lack of physical feedback is also a potentially limiting factor. While we provided haptic feedback to the user via the controllers, there was no force feedback. As the bow is drawn back in the archery task, the controllers vibrate. When the tennis racket hits a ball, it generates a vibration. While these add somewhat to the realism of the experience, they cannot emulate it entirely. That said, our main aim in this paper is to investigate the influence of perspective in dynamic tasks in VR, and it is unlikely that more realistic physical feedback would have a significant systematic influence on the results between our conditions.

Another potential limitation of this study is the generality of the hypotheses. By the very nature of a dynamic task, the rule-set of the task is ever changing. This means that our results may not be applicable to dynamic tasks which do not have similar inputs and outputs as those chosen for this study. For example, a navigational task may require different spatial perception abilities to the tasks here. This is an interesting direction for future work, as we would like to isolate multiple spatial perception and performance DVs which can be empirically analysed to determine how perspective influences each of them, thereby generalising the results to a wide range of dynamic tasks in VR.

Our findings suggest that, with the exceptions of performance in the most dynamic task and spatial perception in the least dynamic task, 1PP is generally better than 3PP for both performance and spatial perception. We also show that A3PP is better than E3PP for performance and spatial perception in most conditions. There is also evidence to suggest that the differences between perspectives in user performance and spatial perception become less pronounced as the task becomes more dynamic. Hence, 1PP is our recommended perspective for static to moderately dynamic tasks in VR, and if 3PP is used, an allocentric version is generally the better choice than egocentric.

Acknowledgements

Naval Bhandari's work on the use and impact of perspective in virtual environments is funded by the EPSRC Centre for Digital Entertainment (grant EP/G037736/1) and BMT Group Ltd. Eamonn O'Neill's research is partly funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications (EP/M023281/1). We thank Crescent Jicol for his help and advice.

REFERENCES

[1] F. M. Alonso, R. Kajastila, T. Takala, M. Matveinen, M. Kytö, and P. Hämäläinen. Virtual ball catching performance in different camera views. In *Proceedings of the 20th International Academic Mindtrek Conference*, AcademicMindtrek '16, pp. 104–112. ACM, New York, NY, USA, 2016. doi: 10.1145/2994310.2994335

[2] S. C. Barathi, D. J. Finnegan, M. Farrow, A. Whaley, P. Heath, J. Buckley, P. W. Dowrick, B. C. Wuensche, J. L. Bilzon, E. O'Neill, et al. Interactive feedforward for improving performance and maintaining intrinsic motivation in VR exergaming. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 408. ACM, 2018.

[3] S. Bateman, A. Doucette, R. Xiao, C. Gutwin, R. L. Mandryk, and A. Cockburn. Effects of view, input device, and track width on video game driving. In *Proceedings of Graphics Interface 2011*, pp. 207–214. Canadian Human-Computer Communications Society, 2011.

[4] J. Bolton, M. Lambert, D. Lirette, and B. Unsworth. Paperdude: A virtual reality cycling exergame. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pp. 475–478. ACM, New York, NY, USA, 2014. doi: 10.1145/2559206.2574827

[5] M. Borges, A. Symington, B. Coltin, T. Smith, and R. Ventura. HTC Vive: Analysis and accuracy improvement. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2610–2615. IEEE, 2018.

[6] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User interfaces: Theory and Practice*. Addison-Wesley, 2004.

[7] B. Brehmer. Dynamic decision making: Human control of complex systems. *Acta psychologica*, 81(3):211–241, 1992.

[8] E. Brunner and F. Langer. Nonparametric analysis of ordered categorical data in designs with longitudinal observations and small sample sizes. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 42(6):663–675, 2000.

[9] T. T. Brunyé, A. Gardony, C. R. Mahoney, and H. A. Taylor. Going to town: Visualized perspectives and navigation through virtual environments. *Computers in Human Behavior*, 28(1):257–266, 2012. doi: 10.1016/j.chb.2011.09.008

[10] I. D. Cherney, C. M. Brabec, and D. V. Runco. Mapping out spatial ability: sex differences in way-finding navigation. *Perceptual and motor skills*, 107(3):747–760, 2008.

[11] M. Claypool and K. Claypool. Perspectives, frame rates and resolutions: it's all in the game. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, pp. 42–49. ACM, 2009.

[12] R. Daneels, S. Malliet, J. Koeman, and W. Ribbens. The enjoyment of shooting games: Exploring the role of perceived realism. *Computers in Human Behavior*, 86:330–336, 2018.

[13] H. G. Debarba, S. Bovet, R. Salomon, O. Blanke, B. Herbelin, and R. Boulic. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one*, 12(12):e0190109, 2017.

[14] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic. Characterizing embodied interaction in first and third person perspective viewpoints. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–72. IEEE, 2015.

[15] A. Denisova and P. Cairns. First person vs. third person perspective in digital games: do player preferences affect immersion? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 145–148. ACM, 2015.

[16] A. Denisova, A. I. Nordin, and P. Cairns. The convergence of player experience questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pp. 33–37. ACM, 2016.

[17] L. Donath, R. Rössler, and O. Faude. Effects of virtual reality training (exergaming) compared to alternative exercise training and passive control on standing balance and functional mobility in healthy community-dwelling seniors: a meta-analytical review. *Sports medicine*, 46(9):1293–1309, 2016.

[18] W. Edwards. Dynamic decision theory and probabilistic information processings. *Human factors*, 4(2):59–74, 1962.

[19] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray. An evaluation of heart rate and electrodermal activity as an objective qoe evaluation method for immersive virtual reality environments. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6. IEEE, 2016.

[20] L. Englund. Using your body to control a 3d game viewed from a 2d third-person perspective. *USCCS 2016*, p. 31.

[21] S. Finkelstein, A. Nickel, Z. Lipps, T. Barnes, Z. Wartell, and E. A. Suma. Astrojumper: Motivating exercise with an immersive virtual reality exergame. *Presence: Teleoperators and Virtual Environments*, 20(1):78–92, 2011.

[22] C. Gonzalez. Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, 46(3):449–460, 2004.

[23] C. Gonzalez. Task workload and cognitive abilities in dynamic decision making. *Human Factors*, 47(1):92–101, 2005.

[24] C. Gonzalez, J. F. Lerch, and C. Lebiere. Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4):591–635, 2003.

[25] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir. First-and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:33, 2017.

[26] D. Henry and T. Furness. Spatial perception in virtual environments: Evaluating an architectural application. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pp. 33–40. IEEE, 1993.

[27] S. K. Hunter. Sex differences and mechanisms of task-specific muscle fatigue. *Exercise and sport sciences reviews*, 37(3):113, 2009.

[28] V. Interrante, B. Ries, J. Lindquist, M. Kaeding, and L. Anderson. Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(2):176–198, 2008.

[29] C. Ioannou, P. Archard, E. O'Neill, and C. Lutteroth. Virtual performance augmentation in an immersive jump & run exergame. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 158. ACM, 2019.

[30] S. Johnson, M. Gibson, and B. Mutlu. Handheld or handsfree?: Remote collaboration via lightweight head-mounted displays and handheld devices. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pp. 1825–1836. ACM, New York, NY, USA, 2015. doi: 10.1145/2675133.2675176

[31] C. Keighrey, R. Flynn, S. Murray, and N. Murray. A qoe evaluation of immersive augmented and virtual reality speech & language assessment applications. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6. IEEE, 2017.

[32] J. H. Kerstholt and J. G. Raaijmakers. Decision making in dynamic task environments. *Decision making: Cognitive models and explanations*, pp. 205–217, 1997.

[33] T. Kosch, R. Boldt, M. Hoppe, P. Knierim, and M. Funk. Exploring the optimal point of view in third person out-of-body experiences. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, p. 68. ACM, 2016.

[34] B. Lang. Analysis of valve's 'lighthouse' tracking system reveals accuracy, 2017.

[35] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. In *Embodiment, ego-space, and action*, pp. 17–60. Psychology Press, 2008.

[36] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa. Collaboration in augmented reality. *Computer Supported Cooperative Work (CSCW)*, 24(6):515–525, 2015.

[37] D. Medeiros, R. K. dos Anjos, D. Mendes, J. M. Pereira, A. Raposo, and J. Jorge. Keep my head on my shoulders!: why third-person is bad for navigation in vr. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, p. 16. ACM, 2018.

[38] D. Monteiro, H.-N. Liang, Y. Zhao, and A. Abel. Comparing event related arousal-valence and focus among different viewing perspectives in vr gaming. In *International Conference on Brain Inspired Cognitive Systems*, pp. 770–779. Springer, 2018.

[39] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017.

[40] D. Peña, M. J. Contreras, P. C. Shih, and J. Santacreu. Solution

strategies as possible explanations of individual and sex differences in a dynamic spatial task. *Acta Psychologica*, 128(1):1–14, 2008.

[41] L. Phillips, B. Ries, M. Kaeding, and V. Interrante. Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 115–1148. IEEE, 2010.

[42] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[43] B. Ries, V. Interrante, M. Kaeding, and L. Phillips. Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. In *Proceedings of the 16th ACM symposium on virtual reality software and technology*, pp. 59–66, 2009.

[44] G. Riva. From virtual to real body: virtual reality as embodied technology. *J. Cyber Ther. Rehabil*, 1:7–22, 2008.

[45] R. Rouse III. What's your perspective? *ACM SIGGRAPH Computer Graphics*, 33(3):9–12, 1999.

[46] P. Salamin, D. Thalmann, and F. Vexo. The benefits of third-person perspective in virtual and augmented reality? In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '06, pp. 27–30. ACM, New York, NY, USA, 2006. doi: 10.1145/1180495.1180502

[47] P. Salamin, D. Thalmann, and F. Vexo. Improved third-person perspective: A solution reducing occlusion of the 3PP? In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '08, pp. 30:1–30:6. ACM, New York, NY, USA, 2008. doi: 10.1145/1477862.1477902

[48] P. Salamin, D. Thalmann, and F. Vexo. Intelligent switch: An algorithm to provide the best third-person perspective in augmented reality. In *Proceedings of the 22nd Annual Conference on Computer Animation and Social Agents (CASA 2009)*, number VRLAB-CONF-2010-001, 2009.

[49] Á. Török, T. P. Nguyen, O. Kolozsvári, R. J. Buchanan, and Z. Nadasdy. Reference frames in virtual spatial navigation are viewpoint dependent. *Frontiers in human neuroscience*, 8:646, 2014.

[50] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang. A subjective study on qoe of 360 video for vr communication. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6. IEEE, 2017.

[51] C. W. VanVoorhis and B. L. Morgan. Understanding power and rules of thumb for determining sample sizes. vol. 3, pp. 43–50, 2007.

[52] M. Vidal, M.-A. Amorim, and A. Berthoz. Navigating in a virtual three-dimensional maze: how do egocentric and allocentric reference frames interact? *Cognitive Brain Research*, 19(3):244–258, 2004.

[53] K. Vogeley, M. May, A. Ritzl, P. Falkai, K. Zilles, and G. R. Fink. Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of cognitive neuroscience*, 16(5):817–827, 2004.

[54] W. S. Waller and T. R. Mitchell. Conditional probability judgments: Effects of imagining vs experiencing the conditioning event. *Organizational Behavior and Human Decision Processes*, 49(2):302–325, 1991.

[55] J. A. Waterworth, E. L. Waterworth, F. Mantovani, and G. Riva. On feeling (the) present. *Journal of Consciousness Studies*, 17(1-2):167–188, 2010.